Stat Papers (2013) 54:545–546
DOI 10.1007/s00362-012-0437-2

BOOK REVIEW

O. Jones, R. Maillardet, and A. Robinson: Introduction to scientific programming and simulation using R

Chapman and Hall, Boca Raton, 2009, xi + 453 pp, US $79.95, £49.99, €65.50 ISBN 978-1-4200-6872-6

Christian Kleiber

Published online: 8 March 2012

During the last decade R has quickly become the industry standard in statistical computing, at least in academia. This is underlined by a rapidly growing number of books on either R itself or on statistical methodology with integrated applications that make use of this environment. However, the vast majority of books on R, or using R, are books on data analysis and hence employ R as a statistical software package, often not going beyond built-in functionality. But R is much more than that: it is a programming language, albeit with a statistical slant. The book under review is one of the few books that propagate R as an environment for computational science. More specifically, the focus is on scientific computing at large as opposed to ‘mere’ statistical computing, with an emphasis on stochastic modeling and simulation. The authors demonstrate that R is well suited for such tasks, offering as it does routines for, inter alia, root finding, numerical integration and differentiation, and also a growing number of functions and add-on packages for optimization and simulation.

As the authors state in their preface, “[o]ur intended audience is those who want to make tools, not just use them.” Overall, the text emphasizes developing programming and modeling skills rather than mere use of canned routines. As has become customary in the R world, there is an accompanying R package, named spuRs, which provides data sets and code from the text.

The book, quite visibly based on course materials, is divided into four parts: Part I, entitled “Programming”, is about basic aspects of the R language, including data structures, scoping rules, object orientation and debugging. Examples include mathematical classics such as the Sieve of Eratosthenes for finding prime numbers. In addition, the authors offer lots of useful advice along the way, for example on good programming habits. Furthermore, there is a chapter on graphics. Part II is about basic numerical

C. Kleiber (✉)
Universität Basel, Basel, Switzerland
e-mail: christian.kleiber@unibas.ch

analysis, beginning with some brief but useful remarks on machine representation of numbers. It moves on to algorithms for root finding and numerical integration as well as optimization in one and several dimensions. In line with the book's aims the authors often prefer to discuss basic algorithms, such as the secant method or Simpson's rule, instead of specialized refinements. Part III covers probability basics, including several statistical distributions, discrete as well as continuous. These chapters are mainly required as background material for Part IV, which deals with simulation techniques. Here, introductory material on the simulation of uniformly distributed random variables, and on the simulation of more general continuous random variables by means of inversion and rejection methods, is followed by chapters on Monte Carlo integration and on variance reduction techniques. A unique feature is a final chapter with case studies. These include the spread of an epidemic, an inventory model and an ecology application on seed dispersal. I particularly liked the various student projects, among them modeling the level of a dam or modeling of insurance risks.

Of course, the material presented in Parts I and III can be found in many sources, in some form or other; it is mainly needed here to keep the book self-contained. The code is usually easy to follow but could have been more compact in some places; the R package also has some minor oddities. Although the text is already somewhat too long for a year-long course (it originated from such a course at the University of Melbourne), I am tempted to suggest a further topic that could serve as a natural bridge between Parts II and IV: heuristic optimization. Basic forms of such algorithms are often not unduly difficult to code, yet they help to illustrate practical uses of simulation beyond statistical inference or the modeling of stochastic systems. Perhaps there is room for this in a future edition.

Overall, the authors have produced a highly readable text. As prerequisites do not go beyond first-year calculus, the book should appeal to a wide audience; it should also be eminently suitable for self-study. On a somewhat larger scale, it may help to further establish R as a kind of Swiss Army knife for computational science. I strongly recommend it.